

10

Approved For Release 2004/01/15 : CIA-RDP64-00046R000200030003-3

With the compliments of  
Ariadne W. Lukjanow  
Rudolf Loewenthal

## REPORT ON:

- 1) SOME PRINCIPLES OF THE UNIFIED TRANSFER SYSTEM (UTS)
- 2) AUTOMATIC DECLENSION OF RUSSIAN NOUNS FOR UTS
- 3) COMPUTER IMPLEMENTATION OF UTS

By

Ariadne Lukjanow  
Rudolf Loewenthal  
B. D. Blickstein

**C-E-I-R** INC

MAIN OFFICE: 734 Fifteenth Street, N.W., Washington 5, D. C.

RESEARCH CENTER: 1200 Jefferson Davis Highway, Arlington 2, Va.

Approved For Release 2004/01/15 : CIA-RDP64-00046R000200030003-3

REPORT ON:

- 1) SOME PRINCIPLES OF THE UNIFIED TRANSFER SYSTEM

By  
Ariadne Lukjanow

- 2) AUTOMATIC DECLENSION OF RUSSIAN NOUNS FOR  
UNIFIED TRANSFER SYSTEM

By  
Rudolf Loewenthal

- 3) COMPUTER IMPLEMENTATION OF UNIFIED TRANSFER  
SYSTEM

By  
B. D. Blickstein

January 1960

C-E-I-R, INC.

Main Office: 734 Fifteenth Street, N.W., Washington 5, D.C.  
Research Center: 1200 Jefferson Davis Highway, Arlington 2, Va.

TABLE OF CONTENTS

Page

SECTION ONE

SOME PRINCIPLES OF THE UNIFIED TRANSFER SYSTEM

By Ariadne Lukjanow..... 1

SECTION TWO

AUTOMATIC DECLENSION OF RUSSIAN NOUNS FOR  
UNIFIED TRANSFER SYSTEM

By Rudolf Loewenthal..... 38

SECTION THREE

COMPUTER IMPLEMENTATION OF UNIFIED TRANSFER SYSTEM

By B. D. Blickstein..... 66

SECTION ONE

REPORT ON SOME PRINCIPLES OF THE UNIFIED TRANSFER SYSTEM (UTS)

By

Ariadne Lukjanow

C-E-I-R, INC.

I. INTRODUCTION

Several approaches have been employed in Machine Translation in the course of the past few years. These approaches were either determined by specific objectives or influenced by the background of the research workers. The objectives range from automatic dictionaries to translations with varying degrees of accuracy, readability, and perfection. The background of a researcher can influence his approach to Machine Translation in three basic ways. One approach may be influenced by machines in such a way that only the development of a new language computer would lead to acceptable results. Another approach may consist of an attempt to simulate human reasoning on a standard computer.

A third approach would be to make Machine Translation as mechanical and utilitarian as possible, by adapting this attempt to the capabilities of the machine and by clearly defining the relationship between man and machine. Since present-day computers are best suited to repetitive mathematical operations and man is still the best thinker, this last approach will make it possible to utilize both of these capabilities to their fullest extent. All thinking will be expressed in the form of codes in the dictionary in the manner provided for by the system.

In order to translate at all, any system must provide solutions to the problem of transferring structure, function, form and meaning from the source language into the target language. Thus, we can call translation a fourfold transfer process consisting of:

- (1) Transfer of the function of words (parts of speech)
- (2) Transfer of the form of words (morphology)
- (3) Transfer of the meaning of words (semantics)
- (4) Transfer of the location of words (syntax)

Every word has a meaning, even if there occurs a so-called "zero-translation," or non-translation. In this system, we shall accept a 1:1 translation as equivalent to no-meaning problem.

Every word in a language has its function; i.e., it is a part of speech and, unless it is a non-translation item, it also has a location or position (syntax) qualification. Transfer process can be visualized as a combination of the following six concepts:

- (1) Function (some "particles," some adverbs)
- (2) Function + location (some punctuation marks, some adverbs, some gerunds)
- (3) Function + form + location (groups from all parts of speech)
- (4) Function + form (some prepositions, some adverbs, some gerunds, negations, etc.)
- (5) Function + form + meaning + location (groups from every part of speech)
- (6) Function + meaning + location (some adverbs, some conjunctions, etc.)

Example:

Combination of function and location:

posle - later; adverb with a 1:1 translation equivalent  
and location "after verb."

Colon, punctuation mark: - 1:1 equivalent, position is at  
the end of a clause.

It is obvious that the elements of the transfer form sets with variants in each of the elements. We can visualize them as follows:

Function	Form	Meaning	Location
x	Ø	Ø	Ø
Ø	x	Ø	Ø
Ø	Ø	x	Ø
Ø	Ø	Ø	x
x	x	Ø	Ø
x	Ø	x	Ø
x	Ø	Ø	x
Ø	x	x	Ø
Ø	x	Ø	x
Ø	Ø	x	x
x	x	x	Ø
x	Ø	x	x
Ø	x	x	x
x	x	Ø	x

Ø - non-variant or absent

x - variant

It would seem that these variations could be expressed in mathematical formulae, but this is not true because the relationship between the variants does not follow the rules of permutation or random combinations. In contrast, these variations follow definite linguistic rules which permit only certain variants within certain combinations. In order to determine these linguistic combinations for the elements of transfer, it is necessary to define and

classify each variant for every element of transfer, as well as the relationship between the variants of each element of the transfer to the variants of the other three.

This can best be illustrated on prepositions:

ELEMENT OF TRANSFER	DEFINITION
function	preposition
form	case government; i.e., prepositions demanding the genitive, dative, accusative, instrumental, or locative
meaning	prepositions of time (static, earlier, later), location or space (where, to where, from where), cause, goal, substitution, division, etc.
location	first item in prepositional phrase, or position 1 in prepositional phrase

Theoretically, we could produce a transfer combination of preposition + dative + location (from where) + position 1 of prepositional phrase, but the grammatical rules and semantic connotations do not permit this type of combination. The prepositions of location are subject to the following division only:

LOCATION	GENITIVE	DATIVE	ACCUSATIVE	INSTRUMENTAL	LOCATIVE
a) where?	bliz vne mezhdu sredi u	po		za mezhdu nad pered pod	v na pri
b) where to?	do	k	v za na pod skvoz6 cherez		o
c) from where?	iz iz-za iz-pod ot s				

The above table shows that the "from where?" definition is used only with the genitive case. Thus, the only usable and meaningful combination is:

preposition + genitive + location (from where?) +  
first position of prepositional phrase

In the UTS we accept any meaningful and valid combination of elements of transfer expressed in the form of numerical digits as a single unified transfer code.

Since many words of the source language can be associated with several function, form, meaning, and location qualifications, it is necessary to combine single transfer code units into sets of codes which can express these variations.

Examples:

dannye      nominal  
modifier

vdol6      preposition of genitive  
adverb

s      preposition of - genitive



sredi	preposition of	location (where?)
		time (static)

If we consider that we have four elements of transfer, each of which has a definite and limited number of variants, it is safe to assume that the number of transfer codes is limited and that we may likewise assume that the same applies to sets of transfer codes. This leads us to the concept that numerous words in the dictionary are associated with identical transfer codes or identical sets of transfer codes. This fact makes possible the concept of code patterns. The number of single transfer code units in the pattern can vary from one to several. After examining some 50,000 canonical entries (stems) in the dictionary of Smirnitskij, we have decided to set the limit at a maximum of 25 single code units in the pattern.

Now let us examine the actual elements of each transfer. Since in translation we are dealing with at least two languages simultaneously, we have to develop a criterion for parts of speech, morphology, semantics, and syntax which would accommodate both languages under consideration, or we must establish a classification system which in form of transfer codes would permit us to place an equal sign between the two languages. This necessitates a certain type of analysis and of synthesis of the grammars of both languages.

## II. THE FUNCTION OF WORDS OR THE CATEGORIZATION OF WORD BEHAVIOR

When examining conventional parts of speech in Russian and English grammars separately, we note that they contain identical categories such as prepositions, adverbs, nominals, modifiers, etc. But when we compare these categories of both languages, we discover that they differ considerably in usage, behavior, and function. In terms of a translation system, this means that either we have to introduce new synthetic categories or we have to divide and redistribute words differently within these categories. Categorizing is, of course, a somewhat subjective process. That can best be illustrated by examining the English preposition "to," in the following manner:

QUALIFICATION	ENGLISH	RUSSIAN EQUIVALENTS	BILINGUAL DATA	TRANSFER DATA (CLASSIFICATION)
Function	1. prepo- sition	1. preposition	1. prepo- sition - like item	1. preposition code
Behavior	2. intro- ducer of infini- tive	2. non- existent	2. particle - like item	2. particle code

Obviously, the second category in the above table might as well be classified as a special auxiliary verb (instead of "particle"), but to the author of the system the definition as "particle" appears more reasonable, perhaps because of the occurrence of the Russian particle "by" in the verbal phrase.

In the process of comparative analysis-synthesis, we have established the following basic categories as transfer parts of speech (listed alphabetically):

- (1) adjectival modifier
- (2) adjective/noun
- (3) adverb (incl. some gerunds and the particle li)

- (4) adverbial modifier (type: bolee, menee, etc.)
- (5) Auxiliary verb (byl, byli, etc.)
- (6) auxiliary verb (moch6, khotet6, etc.)
- (7) conjunction
- (8) negation (incl. some negative adverbs)
- (9) nominal (animate), incl. some pronouns
- (10) nominal (inanimate), incl. some pronouns and numerals.
- (11) nominal (formulae, cardinal numbers, missing words)
- (12) numerical modifier
- (13) particle
- (14) participial modifier
- (15) preposition
- (16) pronominal modifier
- (17) pronoun (type: nami, vami, imi, etc.)
- (18) pronoun (soboj)
- (19) punctuation marks (each treated as a separate category, a total of six)
- (20) verb (including participles such as izucheny, otkryty, etc.)

The assignment of these basic categories to individual words is a discrete and subjective process. It can give valid results only if all other factors and constituent parts of transfer are being taken into consideration. We proceed from the parts of speech as categories to their classification. That can be expressed in the form of a numeric code.

We know that sentences and phrases are combinations of these categories and that these combinations cannot be produced by random distribution of words. Words have to occupy certain positions in order to form a meaningful combination or phrase.

If we take the three-word phrase "in this room," we cannot convey the same idea by a redistribution of the participating words:

"this in room"

"this room in"

"room in this"

"room this in"

"in room this"

We will either get a meaningless jumble of words or convey a different idea. We say "our new building," but not "new our building." We place some adverbs before verbs, some after them. Some of these phenomena can be explained, some are ascribed to usage, but others escape any logical explanation.

Dealing with 26 categories and considering each of them in relation to the other 25, we can establish a hierarchy within the meaningful combinations of parts of speech; i.e., logical sequences.

This point can be illustrated by the position of words within the sequence of a prepositional phrase consisting of a preposition (P), a nominal (N), two adjectival modifiers (AM), and a pronominal modifier (PM):

P before N  
AM before N  
PM before N  
PM before AN  
P before PM  
P before AM  
AM = AM

Thus, we arrive at P-PM-AM-AM-N; or if we assign numerical values to these categories and would like them to form a progression of  $i_1 i_2 i_3$ , etc., we will emerge with the following correlations:

$P < N$ ;  $AM < N$ ;  $PM < N$ ;  $PM < AM$ ;  $P < PM$ ;  $P < AM$ ;  $AM = AM$ ; e.g.,  $P < PM$   $AM = AM < N$ .

Approaching our categories of parts of speech with these criteria, we can assign numerical values or codes to parts of speech (all codes are in octal notation);

01	comma
02	conjunction
16	preposition
17	adverb
20	negation
21	participial modifier
22	pronoun ( <u>nami</u> , <u>vami</u> , etc.)
23	auxiliary verb ( <u>byl</u> , <u>bylo</u> , etc.)
24	auxiliary verb ( <u>moch6</u> , <u>khotet6</u> , etc.)
25	particle
26	verb
27	pronoun ( <u>soboj</u> )
37	adverbial modifier
45	pronominal modifier
46	numerical modifier
47	adjectival modifier
55	adjective/noun
65	nominal (animate)
66	nominal (inanimate)
67	nominal (formulae, numbers, missing words)
70-77	punctuation marks (colon, semicolon, dash, period, etc.)

We are fully aware that this progression method for the identification of a phrase or logical sequence is reliable only in so-called normal sequences. Interrupted sequences or inverted word order require additional re-examination and even actual recognition of constituent parts of sequences. In such cases specific instructions are necessary.

We have, however, established the fact that more than 80%\* of sequences are so-called normal sequences. That frees us of the necessity to recognize at all times every constituent part of all sequences, as well as of every possible combination of the constituent part.

The sequences established through progression codes are by no means permanent or final divisions within the sentence. They can become smaller or disintegrate into single items through either the demands of other components of the transfer process, or through so-called verification instructions.

Example:

A sequence ending with code 47 (adjectival modifier) will call for verification instructions of:

- (1) a sequence within a sequence;
- (2) a sequence with homogeneous parts of speech plus conjunction and/or comma; etc.

We can therefore state that progression codes divide sentences into working units which may or may not become final sequences or phrases. This once more confirms the idea of a total or unified transfer versus a single transfer concept on a different level within the limitations of each phase (structural, morphological, semantic).

The division into sequence is made in accordance with:

$A_1 \dots A_n$  = part of speed code

$B_1 \dots B_n$  = part of speech code with 1st digit being 6

$C_1 \dots C_n$  = part of speech code with value of 10

STOPS

$A_1$		$A_n$
$B_1$		$B_n$ even if $B_1 = B_n$
$A_1$		$C_1 \bar{C}_2 A_n A_1 = A_n$

\* Between 20,000 and 25,000 words in various fields of knowledge have been examined for this purpose.

From: JOURNAL OF CHEMICAL INDUSTRY, vol. 22, no. 9 (1952)

Izucheny reaktsii mezhduehtilovym ehfirom pirokatekhinfosforistoj kisloty i triarilbrommetanami.

Pri vzaimodejstvii ukazannykh soedinenij obrazuiutsia pirokatekhinovye ehfiry triarilmetilfosfinovykh kislot.

Pri omylenii poslednikh slaboj solianoj kislotoj polucheny pirokatekhin i triarilmetilfosfinovykh kisloty.

V nastoiashchem issledovanii nami izuchalis6 reaktsii mezhdue smeshannymi ehfirami fosforistoj kisloty, tipa  $A_2$ ..... i triarilbrommetanami.

Reaktsiia mezhduehtilpirokatekhinovym ehfirom fosforistoj kisloty i triarilbrommetanami po analogii s alkilfosforistymi ehfirami dolzhna idti po reaktsii:  $A_2$ ....

Eksperimental6nye dannye pokazali, chto reaktsiia dejstvitel6no protekaet po ukazannomu uravneniiu.

Tak, naprimer, pri nagrevanii smesi triarilbrommetana i ehtilpirokatekhinovogo ehfira fosforistoj kisloty proiskhodit vydelenie bromistogo ehtila i obrazovanie kristallicheskogo veschestva predstavliaiushchego soboj pirokatekhinovyy ehfir triarilmetilfosfinovoj kisloty.

Dlia ustanovleniia stroeniia poluchennogo soedineniia byla provedena reaktsiia omyleniia razbavlennoj solianoj kislotoj pri nagrevanii ot 180 do 200° v zapalannykh trubkakh.

Produktom omyleniia iavliaiutsia pirokatekhin i triarilmetilfosfinovaiia kislota.

Poluchennyye nami ehfiry tipa  $A_2$ ..... ves6ma ustojchivy k vlage vozdukh.

(As Translated for the Revised English Text)

Concerning the action of triarylbromomethanes on alkylpyrocatechol esters of phosphorous acid.

Reactions between the ethyl ester of pyrocatechol-phosphorous acid and triarylbromomethanes were studied.

(Up)on the interaction of the above-mentioned compounds, pyrocatechol esters of triarylmethylphosphinic acids are formed.

(Up)on hydrolysis of the latter with dilute hydrochloric acid, pyrocatechol and triarylmethylphosphinic acids were obtained.

In the present investigation, the reactions between mixed esters of phosphorous acid of the type ... and triarylbromomethanes were studied /by us/.

The reaction between the ethylpyrocatechol ester of phosphorous acid and triarylbromomethanes should proceed, by analogy with alkylphosphorous esters, according to the reaction: ...

Experimental data showed that the reaction actually proceeds according to the above-mentioned equation.

Thus, for example, upon heating of a mixture of triarylbromomethane and the ethylpyrocatechol ester of phosphorous acid, evolution of ethyl bromide occurs and (there occurs) the formation of a crystalline substance which is the pyrocatechol ester of triarylmethylphosphinic acid.

In order to establish the structure of the compound obtained, a (reaction of) hydrolysis with dilute hydrochloric acid was carried out on heating (at) from 180° to 200° in sealed tubes.

The product(s) of hydrolysis are pyrocatechol and triarylmethylphosphinic acid.

The esters obtained by us of the type ... are extremely resistant to the moisture of the air.



Examining the preceding Russian text sample in terms of the progression code, it appears as follows:

Sentence Number	
1.	26 - 66    16 - 47 - 66    47 - 66    02 - 66 - 77
2.	16 - 66    21 - 66    26 - 47 - 66    47 - 66 - 77
3.	16 - 66    55    47 - 47 - 66    27 - 66    02 - 47 - 66 - 77
4.	16 - 55 - 66    22 - 26 - 66    16 - 21 - 66    47 - 66 01 - 67    67    01 - 66 - 77
5.	66    16 - 47 - 66    47 - 66    02 - 66    16 - 66 16 - 47 - 66    23 - 26    16 - 66 - 70    66    77
6.	47 - 55    26    01 - 02 - 66    17 - 26    16 - 21 - 66 - 77
7.	02 - 01 - 17    01 - 16 - 66    66    66    02 - 47 - 66 47 - 66    26 - 66    47 - 66    02 - 66    47 - 66 - 77 01 - 21 - 27 - 47 - 66    47 - 66    77
8.	16 - 66    66    21 - 66    23 - 26 - 66    66 21 - 47 - 66    16 - 66    16 - 67    16 - 67 16 - 21 - 66 - 77
9.	66    66    26 - 66    02 - 47 - 66 - 77
10.	21 - 22 - 66    67    67    17 - 26    16 - 66 66 - 77

It has been found convenient to make the part of speech code part of the pattern number, so that we can determine the possible logical sequence or wording area immediately after the dictionary look up.

## III. FORM OF WORDS (MORPHOLOGY)

With the part of speech codes, we have devised the means to divide the sentence into possible structural (constituents), sequential (progressive), meaningful combinations, i.e., phrases or fractions of sentence.

The next step would be to establish in which way the constituents of the sequence depend on each other, and what demands they place on each other, if any (i.e., either to confirm the sequence or divide the original sequence into smaller sequences or even single items).

The morphological criteria we are using for this purpose are case, gender, number, and absence of these. For the sake of convenience, we define the demands, government, agreement, influence as "agreement in" case, gender, number.

Numerical values used are:

- (1) Agreement in case # 1 - 7
- (2) Agreement in gender # 1 - 3
- (3) Agreement in number # 1 - 2
- (4) No agreement necessary Ø

Note: Case #7 represents the usage as per example in V

riadu, sadu, na lugu, etc.

V - prep., acc., locative

riadu - nominal in dative

Despite the disagreement in case, it is a meaningful combination in which words "belong together" or form a valid sequence.

Positions of morphological 3-digit code are as follows:

case	gender	number
------	--------	--------

Digits representing case agreement:

- Ø - no case
- 1 - nominative
- 2 - genitive
- 3 - dative
- 4 - accusative
- 5 - instrumental
- 6 - locative
- 7 - auxiliary

Digits representing gender agreement:

- Ø - no gender
- 1 - masculine
- 2 - feminine
- 3 - neuter..

Digits representing number agreement:

- Ø - singular, or no number
- 1 - plural
- 2 - number disagreement (used in impersonal verbs, etc.)

All these morphological qualifications can occur singly or in combinations.

Table of possible morphological codes:

000	010	020	030
100	110	120	130
200	210	220	230
300	310	320	330
400	410	420	430
500	510	520	530
600	610	620	630
700	710	720	730
001	011	021	031
101	111	121	131
201	211	221	231
301	311	321	331
401	411	421	431
501	511	521	531
601	611	621	631
701	711	721	731
002	012	022	032
102	112	122	132
202	212	222	232
302	312	322	332
402	412	422	432
502	512	522	532
602	612	622	632
702	712	722	732

If we consider now the relationships possible between the concepts expressed in part of speech code and the units of morphological codes, we can establish combination sets of codes, i.e., the partial code patterns.

In this report, we shall do so for one part of speech -- the prepositions.

Prepositions, as we know, do not demand "agreement in" number or gender. Therefore, we are dealing with only a case agreement.

Prepositions table:

CASES	ONE CASE PREP.	2 CASE PREP.	3 CASE PREP.	2 CASE & AUXILIARY
2 (genitive)	bez, bliz, vdol', vmesto, vne, vnutri, vozle, vokrug, dlia, do, iz, iz-za, iz-pod, krome, mimo, nakanune, okolo, ot, posle, posredi, protiv, radi, sredi, u	mezhdu (mezhd)	s	
3 (dative)	k, blagodaria, vopreki, podobno, soglasno, naperekor, navstrechu		po	
4 (accusative)	pro, skvoz', cherez	v, na, za, pod, o(ob)	s, po	v, na
5 (instr.)	nad, pered	za, pod, mezhdu (mezhd)	s	
6 (locative)	pri	v, na, o(ob)	po	v, na
7 (auxiliary)				v, na

On the basis of this table, we can say that some prepositions (code 16 or 17) can be associated with one, two, or three morphological units.

The total code patterns will be then as follows:

ONE UNIT PATTERNS	2 UNIT PATTERNS	3 UNIT PATTERNS
16 - 200	16 - 200 - 500	16 - 200 - 400 - 500
	16 - 400 - 600	16 - 300 - 400 - 600
16 - 300	16 - 400 - 500	16 - 400 - 600 - 700
	17 - 200 - 000	
16 - 400	17 - 300 - 000	
	17 - 400 - 000	
16 - 500		
16 - 600		

In this fashion, the 49 prepositions of the prepositions table are associated with 14 code patterns which would accomplish function and form transfer.

## IV. MEANING OF WORDS AND MEANING CLASSES

The analysis of languages in establishing meaning categories is of subjective character and is based on a mental process, which not only requires an intimate knowledge of the languages to be analyzed but also a very careful manipulation of the numbering system in order to prevent an unintentional conflict of meaning classes in the code patterns within sequences.

In this report, we shall attempt to establish some of the criteria of analysis and the nature of classification of semantic or meaning definitions.

In this area we have to make a distinction between situations which can be described as "a word by itself" and "a word in different environments." There are distinctly two levels of meaning ambiguity: (1) a so-called subject matter ambiguity which fits into the category of "a word by itself," and (2) an environmental ambiguity, i.e., "a word in different environments."

Example of the first type of ambiguity:

AKT	=	1. act 2. legal deed (law) 3. convocation (education)
OBRAZOVANIE	=	1. education (education) 2. formation (technical subject matter)
Board	=	1. piece of wood 2. food (household arrangement) 3. stage (theater) 4. a council (political science) 5. an action, as in "to board a train"

The subject-matter ambiguity can be solved sometimes through the environment; for instance, if we encounter the word obrazovanie with modifiers like Kristallicheskie (crystal), Kislotnoe (acid), etc., there is no doubt that the meaning of this word is formation. Outside of environmental influences, we have to depend on the subject matter of the article or book to be translated, i.e., microglossary, and use that as a cue for selection. We have to expect on the level of subject matter meaning ambiguity a certain degree of

The second level of meaning ambiguity, the environmental ambiguity, is subject to meaning categorization or classes.

We shall describe the method of arriving at these classes, as well as some class definitions, through the examination of environment relationships in Preposition - Nominal sequences.

Prepositions by their meaning connotations can be divided into a variety of groups. We shall list some of them here:

(1) Prepositions of time

Simultaneous/Static (when)	Earlier ("before when")	Later ("after when")
v, za, na, po, pri, s, sredi	do, k, za, pered	ot, po, s, cherez

(2) Prepositions of space and location

Where	Where to	Where from
bliz, v, vne, za, na, nad, mezhdu (mezh), pered, po, pod, pri, sredi, u	do, v, k, za, na, o (ob), pod, skvoz', cherez	iz, iz-za, iz-pod, at, s

(3) Prepositions of cause

For whom, for what, why, etc.
za, iz, iz-za, ot, po, s

(4) Prepositions of goal

dlia, do, v, za, k, na, po, radi
----------------------------------



(5) Part of the whole

iz, po

(6) Exchange or replacement

za, vmesto

(7) What is it made from

iz

Now let us consider the category of space or location connotation. We have already divided this category into three sub-categories: (1) The first sub-category implies the specific position of something, generally recognized by yielding an answer to the question, "Where?" It implies a point of location. (2) The second sub-category implies the concept of something proceeding towards a certain location, generally recognized by yielding the answer to the question, "Where to?" It implies a point of destination. (3) The third sub-category expresses the idea of something coming from a certain location, generally recognized by yielding the answer to the question, "From where?" It implies a point of origin.

These sub-categories in turn can be divided further by analyzing specific prepositions.

The prepositions s, iz, and iz-pod, all belong to the "From where?" or point of origin class. They differ in their semantic content on an individual basis. When s is used, it designates either the place from which the object is removed by some agent, or the place from which an object capable of locomotion is leaving. This last instance usually

involves geographic locations in connection with persons or modes of transportation of persons. When iz is used, it designates an object leaving by any means any location that has an egress, or the emergence of an object from another object. When iz-pod is used, it designates an emergence in any manner from under something on the part of an object.

Examples: s Kavkaza, s gory, s sobrania, etc.  
 iz goroda, iz derevni, iz avtomobiia, etc.  
 iz-pod kamnia, iz-pod stola, iz-pod knigi, etc.

The prepositions do, k, cherez, all belonging to "Where to?" or point of destination sub-category, again differ in their meaning content on an individual basis. When do is used, it designates the direction of movement with the definite connotation of limitation or boundary. When K is used, it again designates the direction of the movement, but its definite connotation is to achieve only proximity to the destination. When cherez is used, it designates a penetrating movement through some medium, usually with some difficulty attached to it, and it also designates a movement of directly surmounting a difficult medium.

Examples: do goroda, do Washingtona, etc.  
 K beregu, K gorodu, K reke, etc.  
 cherez les, cherez bar'er, etc.

The prepositions u, bliz', and pri, all belonging to the "Where?" or specific location sub-category, differ on an individual basis in their semantic content. When pri is used, it designates that one object is adjoining another one. On the other hand, u indicates immediate closeness of objects; bliz', in turn, indicates only closeness of objects.

Examples: u reki, u berega,  
 bliz' goroda, pri stantsiii, etc.

In most of these instances, the translation of the prepositions is at variance with their literal meaning (1:1 equivalent).

	PREPOSITIONS	LITERAL MEANING	SPECIAL MEANING
"FROM WHERE" CLASS	s iz iz-pod	with from from under	from from from under
"TO WHERE" CLASS	do k cherez	to to through	to toward through/over
"WHERE" CLASS	u bliz' pri	at near at	by near at

All these categories/classes in turn, have to be divided again into smaller groups. For example: iz-pod in relation to location-objects does definitely mean from-under, but with location-cities the meaning of it becomes from the vicinity of.

Example: iz-pod stola = from under the table  
iz-pod Washingtona = from the vicinity of Washington

The prepositions so far have been analyzed for their special or locational relationships. The same prepositions can also be analyzed with the view of other semantic criteria.

For example the preposition iz with the connotation of selection will in some instances keep the translation from, but in the environment of

- a) before plural pronouns: nikh, vsekh, nas, tekh, etc.
- b) before numerals: dvukh, trekh, etc.
- c) before collective nominals like: chlenov, predstavitelej, iuristov, etc.

will become the preposition of selection, that is, one of many or part of the whole with the translation of.

Example: luchshij iz vsekh = best of all

Komitet iz predstavitelej = committee of representatives

Let us follow through the analysis of the same prepositions with the view of a time relation concept. In this case we will find that three sub-categories become apparent: (1) The first one implies that an action or state of being occurs after a fixed time span. These prepositions are: ot, s, cheraz, po. (2) The second sub-category connotes that an action or state of being occurs before a fixed time span. These prepositions are: do, k, iz, pered. (3) The third one infers that the action or state of being occurs during a fixed time span. These prepositions are: sredi, po, v, s, za, na, pri. Therefore, we can now draw an analogy with the three previously examined sub-classes and can call these time sub-categories "after when," "before when," and "when."

It becomes apparent at this time that some prepositions occurring in these new sub-categories have participated in the previous ones.

Of special interest in analyzing these prepositions are some that coincide with the sub-classes which were previously established:

	PREPOSITIONS	LITERAL MEANING	WHERE CLASS	WHEN CLASS
FROM-AFTER	s	with	from	since
TO-BEFORE	k	to	toward	toward
WHERE-WHEN	pri	at	at	during

For large-scale translation it is necessary not only to apply these larger categories and their sub- and sub-sub-categories, but also to analyze them in terms of each other in order to establish the similarities, as well

as conflicts, and then establish final definite categories. This task has been accomplished in the Unified Transfer System, and the precise description of each category will be included in the projected Unified Transfer System Manual. In this report, due to the limitations of time and the size of the report, we shall limit ourselves to the method of arriving at the categories, rather than categorization itself.

Let us continue the analysis of the three prepositions s, k, and pri, limiting ourselves to "where" and "when" categories.

From initial inspection it would appear necessary to assign at least ten meaning classes, one for each concept of the three prepositions. It becomes apparent that the number of meaning classes then would increase geometrically with the increasing number of preposition-participants and/or the introduction of new categories. We, therefore, begin to search for means of reducing the progression. The cues for this reduction come from two basic sources, and we can estimate that morphology provides about 70% of the cues by imposing the case restrictions, and the usage of language should provide the remaining 30%.

In establishing meaning classes, we can combine several concepts into a single class, as long as there can be no conflict at the morphological or usage levels. The idea is to bring together the three elements of transfer (function, form, and meaning) and reduce the number of code patterns to a minimum. Thus, the next step would be to establish what part morphology plays in reducing the number of meaning classes in the sample prepositions.

This can be illustrated in the following table.

PREPOSITION	CLASS	CASES				
		Genitive	Dative	Accusative	Instrumental	Locative
s	1	from		for	with	
	2	from				
	3	since				
k	1	to				
	2	toward				
	3	toward				
pri	1					at
	2					at
	3					during

Note: 1 = literal; 2 = "where" case; 3 = "when" case.

From the table above we can see that for the preposition s only one morphological category is affected; therefore, we need not assign these particular meaning classes for the accusative and instrumental cases and thus achieve a reduction from the original possible ten to six meaning classes.

Next we note that the meaning classes for each preposition of the sample belong into separate cases, and this would permit us to assign only two meaning classes, a "where" and a "when" class, coded up as entries in the proper case for the respective preposition.

If we now examine these sample prepositions further, we will find that apparently in the case of the preposition s we would not be able to reduce the number of meaning classes any further by virtue of the fact that the "where" translation differs from the "when" translation. On the other hand, k would not present any problem since both translations are identical, while pri again presents us with separate translations for the "where" and "when" cases, but the identical translation of the literal meaning and the

"where" case would permit us to eliminate the "where" class in this instance. Sometimes, when the translation of prepositions with the same meaning definition changes several times, we found it practical to give the preposition a so-called zero translation and attach the translation of the preposition to the nominals. Thus, we again achieve a reduction in the number of meaning classes. The same type of classification is applied to the remaining members of the prepositional phrase (nominals, modifiers) and the classes are assigned within the boundaries of function-form criteria. The definitions of meaning have to be used carefully and discretely, since the criterion of time element is not necessarily derived from strict time terms, but can be arrived at in combinations of prepositions with nominals of action.

Example: pri okislenii = during the oxidation  
pri rabote = during the work  
pri issledovanii = during the investigation

When assigning these classes to nouns it becomes apparent that these concepts are not rigid rules but are the result of subjective judgment. It is impossible, for instance, to say that all nouns of action in the locative case will yield the same translation during for the preposition pri. There are vagaries of usage which defy any definition. For instance, if zhelanii, a noun of action in the locative (desire, wish) occurs with the preposition pri, the translation of the preposition changes to if and the meaning connotation from time to condition.

Example: pri zhelanii = if desired

A final consideration in assigning meaning classes must be an expression of the physical location of the participating members (words), whether they must be immediately adjacent or whether they can be separated by non-participating words, i.e., whether the preposition-nominal relationship is dependent on their immediately adjacent physical location.

Example:

pri tshchatel'nom issledovanii = during careful investigation  
 pri issledovanii = during investigation

versus:

na drugoi den6 = next day  
 na den6 = for a day

Thus, the indication of the boundary, known in the system as boundary indicator or item count, is introduced together with some of the meaning classes. If no boundary is necessary, this indicator is coded as zero, otherwise it corresponds to the number of participating items, e.g., four participant-members (words) require digit 4 as a boundary indicator.

Thus, after complete analysis, the code pattern for preposition with:

- a) one literal translation
- b) P + N = Adverb

would look as follows:

Pattern #1 = 16-0112-200 = Ø translation (adv. class)

-0111-200 = literal meaning (1:1 equivalent)

This brings together the Function-meaning-Form transfer categories.

Example:

Preposition <u>bez</u>	Nominals: interesa
with literal meaning	somneniia, etc.
<u>without</u>	
	pol'zy
	nuzhdy, etc.

bez interesa = without interest  
 bez somneniia = without doubt  
 bez pol'zy = useless  
 bez nuzhdy = needless



## V. LOCATION, ARRANGEMENT OR SYNTAX

The next step in translation, once the function-form-meaning transfer has been achieved in the code, is the transfer of structure, which we could visualize as consisting of two sub-transfers:

- (1) actual structure transfer;
- (2) pure relocation of items within the structural boundaries.

The structure sub-transfer is the division of a total sentence into "sentences" and/or clauses; clauses into blocks; and blocks into phrases.

The first phase of this sub-transfer is the identification of all punctuation marks within the sentence in regard to their meaning and function within the sentence.

For example, let us examine the possible meaning and functions of the semi-colon (;).

Its position:

- (1) between independent "sentences," combinations of which form the total sentence without use of conjunctions. These "sentences" can have commas inside themselves.
- (2) between independent sentences, which are combined into a total sentence by means of
  - (a) conjunctions no, odnako, vse zhe, tem ne menee, etc.
  - (b) conjunctions i, da
- (3) between phrase-type homogeneous members of the sentence, specifically where these "phrases" have modifiers or modifier-groups separated by commas.
- (4) between several subordinate clauses with one main clause present in the sentence; in that case, however, it would not be followed by conjunctions.

(5) between "sentences" which consist of main clause and subordinate clauses, i.e., independent "sentences".

(6) between enumeration or recapitulation.

If we examine the above we can see that with the exception of (3) and (6), the semi-colon (;) is always a division mark between "sentences" and can be, upon recognition on the basis of its function code (71), considered as a stop signal between "sentences" in structural analysis.

For identification of the case in point (6), we have to locate the colon. Then the situation would be:

xxxxxxx:	xxx; xxxx; xxx; xxxxx.	<u>Note:</u> x = word
"sentence"	sentences or phrases	

For identification of the case in point (3), the situation is either similar to point (6) (e.g., we will locate the colon), or if the colon is absent, we are dealing with homogeneous phrases which can be treated in the same fashion as point (1).

Therefore, in the Unified Transfer System the semicolon is accepted as a stop signal for the division of a sentence into clauses or "sentences."

The same process is applied to other punctuation marks until we divide the sentence into "sentences," which then in turn have to be divided into introduction, subject, predicate and final blocks, in the order of their occurrence in a sentence. Then these blocks, which are found to be present, are rearranged into a model structure of introduction-subject-predicate-final blocks.

For the sake of the discussion, let us consider the identification of the subject.

A subject can be:

- (1) Any part of speech in the nominative
- (2) Combination of words (cluster) with the connotation of a

"jointness" like: my s tobol, sestra s bratom, etc

- (3) Numerical combinations with a precise or approximate definition of objects, like: dva priiatelia, neskolko chisel; minogo liudej, etc.

Therefore, we could say that subject can be:

<u>SUBJECT</u>	<u>EXAMPLES</u>
1. Noun	1. <u>Kolba</u> stoiala na stole
2. Adjective	2. <u>Serye</u> ne podlezhata analizu
3. Participle	3. <u>Spavshie</u> prosnulis'
4. Numeral	4. V vode rastvorilis' tolko <u>dva</u>
5. Pronoun	5. Ne poshel <u>on</u> domoj
6. Verb in infinitive	6. <u>Pisat'</u> trudno
7. Non-inflected word	7. Gromkoe " <u>ura</u> " narushilo tishinu.
8. Word cluster	8. <u>Brat s sestroy</u> uchat'sia v universitete
9. Numerical combination	9. <u>Dvesti studentov</u> izuchaliut inostrannye iazyki.

By examining the "sentence" for the presence of a noun in nominative any other part of speech in nominative, which can be then either subject itself or modifier to the subject in other case (i.e., cluster subject), words like mnogo, malo, etc., and identifying it on the basis of their function-form code we can identify the subject and its or their modifiers and rearrange the total into a block sequence.

Example:

Predicate(2)	Subject(1)	final(3)	Original block order (inverted word order)
Ne poshel	on	domoj	
Subject(1)	Predicate(2)	final(3)	Rearranged block order
on	ne poshel	domoi	

Whenever the blocks are larger than one word (which happens in most of the cases), we check the rearrangement numeric codes, which are attached to the translation. These numbers are assigned on the basis of the target language translations as well as the function of its source equivalent.

Example: Dlia naibolee nagliadnogo predstavleniia = to get a clearer concept.

The source language preposition dlia received a verbal translation, but retained a prepositional rearrangement code because of source language word function.

A list of rearrangement codes and their definitions or equivalents follows:

Syntax codes:

- 001 Preposition
- 002 Introductory words (if, that, which, what, how, why, as, since, etc.)
- 003 Conjunctions (and, but, either, or, neither, nor, comma, colon, semicolon, etc.)
- 004 Words like only, just, then, perhaps, maybe, therefore, however, almost, likewise, etc.
- 005 Not
- 006 It is possible, it is not possible, it is known, etc.
- 007 Some, all, any, none, something, anything, any kind, nothing, etc.
- 011 My, yours, his, etc. (possessive form)
- 012 Numerals (one, two, etc., first, second, etc., few, many, much, more, most, last, etc.)
- 013 Other
- 014 Adjectives (including some pronouns)
- 015 Nouns and pronouns (nominals)
- 016 Myself, yourself, itself, etc.
- 017 Participial modifiers
- 021 Will, may, must, can, do, etc.

- 022 Have
- 023 To be
- 024 Seldom, often, really, verbally, continuous, ever, never (adverbs of time)
- 025 Verbs and short forms of participles, adjectives
- 026 Here, there, away, beyond, upstairs (place)
- 027 Equally, rapidly, strangely, unequally, vastly, greatly, considerably, quite, etc. (manner)
- 031 Early, late, later, soon, etc. (time)
- 032 Period

On the basis of the above table, the constituent parts of the system are as follows:

1. Source language dictionary (in our case Russian Dictionary).

Its format is: Russian word = Dict. Line # + Code Pattern #.

Example:

PREDSTAVLENIIA = 14155* 66-212
--------------------------------

\* arbitrary D. L. #

2. Target language dictionary, in our case English dictionary.

Its format is: Dictionary Line # - English equivalent + rearrangement code.

Example:

14155 A	concept	015
B	concept	015
C	of the performance	015
D	the performance	015
E	performance	015
F	of the concept	015
G	the concept	015
H	the performances	015
J	performances	015
K	the performances	015
L	performances	015

## 3. Code patterns arranged by code pattern #.

Example:

66-212	4115 230 000
	3116 230 000
	0111 230 421
	0111 230 401
	0111 230 400
	0111 230 021
	0111 230 001
	0111 230 000
	0111 131 401
	0111 130 400
	0111 431 401
	0111 431 400

The code distribution in pattern:

66-212            66 is functional part of speech

212 actual # of the pattern

line of 10 digit code, unified transfer code:

1	2	3	4	5	6
1 digit	3 digits	3 digits	1 digit	1 digit	1 digit

1. Semantic (meaning) boundary indicator
2. Semantic class
3. Form or morphology group
4. Subject matter indicator (microglossary control)
5. Preposition control
6. Article control

The system instructions outside of strict data preparation (text preparation, dictionary look-up, sorts, etc.) are divided as follows:

1. Progression instructions (determination of working area)
2. Selection:
  - a) comparison of codes in patterns for the selection of form and meaning
  - b) article, preposition selection

3. Verification or correction instructions for:
  - a) changes in progression (larger or smaller string, single word selection etc.).
  - b) Pronoun, conjunction selection
4. Sentence recognition and division instructions
5. Syntactic block recognition
6. Rearrangement instructions

LIBRARY OF CONGRESS

UTS

A  
B  
V  
G  
D  
E  
ZH  
Z  
I  
✓  
I  
K  
L  
M  
N  
O  
P  
R  
S  
T  
U  
F  
KH  
TS  
CH  
SH  
SHCH  
"  
Y  
'  
✓  
E  
IU  
IA

A  
B  
V  
G  
D  
E  
ZH  
Z  
I  
J  
K  
L  
M  
N  
O  
P  
R  
S  
T  
U  
F  
KH  
TS  
CH  
SH  
SC  
'  
Y  
6  
EH  
IU  
IA



SECTION TWO

REPORT ON THE AUTOMATIC DECLENSION OF RUSSIAN NOUNS  
FOR THE UNIFIED TRANSFER SYSTEM (UTS)

By

Rudolf Loewenthal

Key to Abbreviations

a or anim.	animate
acc.	accusative
b	both (animate or inanimate)
C	class
dat.	dative
F or fem.	feminine
gen.	genitive
i or inan.	inanimate
inst.	instrumental
loc.	locative
M or masc.	masculine
N or neut.	neuter
P or plur.	plural
S or sing.	singular

## Numbers in declension patterns;

- 1 - nominative
- 2 - genitive
- 3 - dative
- 4 - accusative
- 5 - instrumental
- 6 - locative

## I. INTRODUCTION

Our transliteration from the Russian follows basically that of the Library of Congress. We omitted all diacritical marks because they cannot be reproduced on the computer. In addition, we made a few alterations; for reasons of economy we reduced the four-letter combination to two English letters corresponding to a single Russian letter. Thus, for practical considerations imposed upon us by the use of a computer, we introduced the following minor changes in the transliteration of the Library of Congress:

Russian	Library of Congress	UTS transliteration
й	й	j
щ	shch	sc
ъ	"	'
б	'	6
э	e	eh

A key problem in the development of the UTS or of any other system is the preparation of a dictionary. The task is twofold:

- (1) The compilation of all the paradigmatic forms of nouns, adjectives, pronouns, verbs, and all verbal forms like participles, gerunds, etc., in the source language; and
- (2) the meanings in the target language.

We are here concerned with the first; that is, the tedious, time consuming, and costly mechanical task of compiling a paradigmatic dictionary manually.

We realized the magnitude of that task, which would have kept us occupied for several months, even if a competent linguistic staff had been provided. The cost would have been high and our energy would have been drained away in organizing and revising such work.

It has been of great help to us that statistical research had been done on some of our problems. Harry H. Josselson of Wayne University made a superb study on "The Russian Word Count,"\* in which he analyzed almost 47,000 Russian words of representative sources for their usage. He came to the conclusion that only 4.4% of the words are non-inflected, i.e., adverbs, prepositions, conjunctions, and participles. 95.6% of all the words are inflected; that is, they are either declined or conjugated.

The findings are as follows:

Group 1:	55.1%	noun, adjective, adjective used as noun, pronoun, and numerals characterized by declension (declensions: 69%)
Group 2:	26.6%	verb
Group 3:	<u>13.9%</u>	verb derived forms of participle, participle used as adjective, participle used as noun, and gerund
Total	<u>95.6%</u>	

After we have established the fact that 95.6% of Russian words in average texts are inflected, let us examine the details of the problem. The Russian language has a total of 137 suffixes for declensions and conjugations, including the suffix Ø which gives the machine negative information.

\* The Russian Word Count And Frequency Analysis Of Grammatical Categories Of Standard Literary Russian. Detroit, Wayne Univ. Press, 1953, p. 12 & 17.

On the average, each of these suffixes occurs close to 2,900 times in a paradigmatic dictionary, based on a 50,000 word dictionary. This should once for all dispel the concept of solving Machine Translation with the help of a split dictionary. The ambiguity in the stem-suffix relationship would simply be staggering. As an example in which it is impossible to identify different parts of speech or grammatical categories, let us examine the stem smotr- (inspection (noun); to look (verb))

noun failure:	-u noun stem	(dative; to inspection)
	-u verb stem	(non-existent)
verb failure:	-iu noun suffix	(non-existent)
	-iu verb suffix	(I am looking)

We calculated that approximately one in seven stems (one sixth), or almost 17%, is subject to such ambiguity on the basis of the 60,000-word dictionary by Miller.

The magnitude of the process of compiling a paradigmatic dictionary can be gauged with fair accuracy. Noun or adjective stems have usually ten paradigms, while verbs have 37 paradigms. Russian stems average 25 paradigms, including Ø. Thus for a dictionary of 50,000 terms, it would be necessary to compile a paradigmatic dictionary of some 1,250,000 paradigmatic forms. We have found a solution to this problem by utilizing an electronic computer.

Traditionally, there are three declensions and two (or possibly three) conjugations. By substantially increasing the number of categories we were able to eliminate 95% of the irregularities, thus leaving us with less than 5% of the words (approx. 2,500) to be declined or conjugated manually. For use on the computer, we have assigned each of our declension or conjugation categories a code number which we call a "declension code pattern." In accordance with the code pattern numbers, the computer will

select appropriate suffixes for every stem entered and compile a complete set of paradigms for every word desired.\* We can increase our dictionary by a minimum of 5,000 words or 125,000 paradigmatic forms per month and provide them with all the codes necessary for translation.

---

\* We are using stems with a maximum of invariable letters.

II. DECLENSION OF NOUNSA. General

For machine purposes, we have established 11 declension groups with 19 sub-groups, or altogether 30 classes of Russian nouns. Only two of the neuter groups are quite regular and have no sub-groups (see Chart 1 at the end of this section, nos. 3 and 11). The frequent masculine (nos. 1 and 2) and neuter (nos. 3 and 4) declensions have been placed next to each other at the beginning of the chart because of their many similarities. The minor groups (masc.: nos. 8, 9, and 10; neut.: no. 11) appear at the end of the list. In the feminine gender (nos. 5, 6, and 7; incl. two small masc. sub-groups: nos. 5c and 6d), the distribution is more even. The main declension patterns (nos. 1-7) with their sub-groups apply to between 80% and 85% of all the nouns, while the minor patterns (nos. 8-11) provide for about 10% of the remaining nouns. Approximately 5% of the nouns are quite irregular and will have to be prepared by hand.

There are only two variant forms for the dative, instrumental, and locative plural of all nouns. In contrast, the endings of the other cases have numerous variations.

B. Undeclinable Nouns\*

Certain types of nouns cannot be declined at all. They include:

- (1) Foreign names or loan words ending in -i and -u/-iu (alibi, Peru, parveniu).
- (2) European loan words ending in -o and -e/-eh (kakoa, kofe, aloeh).
- (3) Ukrainian surnames ending in -ko.
- (4) Some foreign loan words or names ending in a stressed -a (amplua, Diuma).
- (5) Feminine nouns (applying to women) ending in a consonant: madam, miss, missis, frejlejn, mademuazel6; likewise alma mater and bersez.

\* Undeclinable nouns can be traced in any standard Russian grammar. The above information was summarized from the useful presentation of B. O. Unbegaun in his Grammaire russe (Paris, Les Langues du Monde, 1953, p. 70-71).

- (6) Surnames ending in a consonant, even if they are of Russian or Ukrainian origin (volk, Gogol6).
- (7) Russian surnames ending in -ykh/-ikh, -ogo, and -ago.
- (8) Musical notes (do, re, mi, fa, so, la, si) and the letters of the alphabet.

### C. Declension of Masculine Nouns

There exist altogether 17 automatic declension patterns for the masculine gender; i.e., two main groups and three minor groups; the remainder are sub-groups (see table 1).

TABLE 1: SAMPLES OF 17 MASCULINE NOUN DECLENSION PATTERNS

Number	Number on Chart	Number of Declension Pattern	Sample Word	Animate or Inanimate
1	1	1a	doklad	i
2	2	1b	ocherk	i
3	3	2a	saraj	i
4	4	2b	delitel6	i
5	5	2c	put6	i
6	6	2d	ehkipazh	i
7	7	2e	geroj	a
8	22	8a	uchenik	a
9	23	8b	lev (2 stems)	a
10	24	9a	kon6	a
11	25	9b	uchitel6	a
12	26	10a	gorod	i
13	27	10b	nozh	i
14	28	10c	tovarisc	a
15	29	10d	doktor	a
16	15	5c	muzhchina (fem. pattern)	a
17	Approved For Release 2004/01/15 : CIA-RDP64-00046R000200030003-3			a



Masculine nouns display the greatest variety of declension patterns, actually they account for 17 or more than one-half of the total.

M-1 This very large class of words pertains to inanimate objects.

Table 2 - Class 1

C	S		P	
	la(i)	lb(i)	la(i)	lb(i)
1	∅	∅	y	i
2	a	ov	ov	ov
3	u	u	am	am
4	∅	∅	y	i
5	om	om	ami	ami
6	e	e	akh	akh

Samples: la: vint, narod, doklad, pantograf, son (2 stems)

lb: ocherk, potok, ugolek (2 stems), veter (2 stems)

Sub-group lb differs from la only in the ending -i

(instead of -y) in the nom. and acc. plur.\*

---

\* There are many masc. nouns ending in a compound consonant. Their gen. plur. usually ends in -ej (10b-c).

M-2 This major group has the following endings:

Table 3 - Class 2

C	S					P				
	2a(i)	2b(i)	2c(i)	2d(i)	2e(a)	2a(i)	2b(i)	2c(i)	2d(i)	2e(a)
1	j	6	6	ø	j	i	i	i	i	i
2	ia	ia	i	a	ia	ev	ej	ej	ej	ej
3	iu	iu	i	u	iu	iam	iam	iam	am	iam
4	j	6	6	ø	ia	i	i	i	i	i
5	em	em	em	em	em	iami	iami	iami	ami	iami
6	e	e	i	e	e	iakh	iakh	iakh	akh	iakh

Samples, 2a: sluchaj, saraj, kraj, boj, aluminij (sing. only),  
chaj (sing. only)

2b: delitel6, koren6, den6 (2 stems), kamen6, slovar6

2c: (rare): put6

2d: (extremely rare): ehkipazh

2e: geroj, vorobej, muravej, sviditel6

Sub-group 2e is the only class of animate nouns in this major declension group and corresponds directly to sub-group 2a.

M-8 This minor class includes only animate nouns:

Table 4 - Class 8

C	S		P	
	8a(a)	8b(a)	8a(a)	8b(a)
1	ø	ø	i	y
2	a	a	ov	ov
3	u	u	am	am
4	a	a	ov	ov
5	om	om	ami	ami
6	e	e	akh	akh

Samples, 8a: letchik, starik, byk, brat, voron, izvozchik, volk

8b: lev (2 stems), ugol (2 stems), vol

Sub-group 8b differs from 8a only in the nom. plur., where the two-stem nouns carry the ending -y (instead of -i). Masc. nouns of class 8a ending in the letters -k, -g, -kh, -ch, -sc, -zh, -sh carry the ending -i in the nom. plur. The inanimate nouns of classes 10b (very close to 1b) and 10c have the -i ending also in the acc. plur.

M-9 The endings of this minor group of animate nouns are as follows;

Table 5

- Class 9

C	S		P	
	9a(a)	9b(a)	9a(a)	9b(a)
1	6	6	i	ia
2	ia	ia	ej	ej
3	iu	iu	iam	iam
4	ia	ia	ej	ej
5	em	em	iam1	iam1
6	e	e	iakh	iakh

Samples, 9a: kon6, golub6, spasitel6, pisatel6, zhitel6, gost6

9b: uchitel6, rukovoditel6.

The very small 9b class has the ending -ia in the nom. plur. (instead of -i; cf. also classes 10a and 10d).

M-10 This is one of the most ambiguous classes, containing animate as well as inanimate nouns.

Table 6 - Class 10

C	S				P			
	10a(i)	10b(i)	10c(a)	10d(a)	10a(i)	10b(i)	10c(a)	10d(a)
1	ø	ø	ø	ø	a	i	i	a
2	a	a	a	a	ov	ej	ej	ov
3	u	u	u	u	am	am	am	am
4	ø	ø	a	a	a	i	ej	ov
5	om	om	em	om	ami	ami	ami	ami
6	e	e	e	e	akh	akh	akh	akh

Samples: 10a: gorod, ostrov, porus, vecher

10b: nozh

10c: tovarisc

10d: doktor

Differences between the various groups occur in the acc. and inst. sing., as well as in the nom., gen., and acc. plur. Note the unusual plur. ending -a in the nom. (10a and 10d) and acc. (10a) plur.

M-5c, M-6d. Two small sub-groups of masc. nouns are declined like fem. nouns and have been classified among them. Their endings are:

Table 7

Classes 5c and 6d

C	S		P	
	5c(a)	6d(a)	5c(a)	6d(a)
1	a	ia	i	i
2	i	i	ej	ej
3	e	e	am	iam
4	u	iu	ej	ej
5	ej	ej	ami	iami
6	e	e	akh	iakh

Samples, 5c: muzhchina, dedushka, mal6chishka

6d: diadia, sud6ia

D. Declension of Neuter Nouns

For automatic declension, we consider all neuter nouns as inanimate. The word ditia, though animate, is so irregular that it requires manual declension. As a matter of fact, most of the irregular nouns, which have to be declined by hand, belong to the neuter nouns.\* Other rare exceptions of neuter animate nouns we integrate into one of the masculine classes for our purposes.

The six neuter declension patterns (classes 3, 4, and 11) are much more regular than the masculine ones. Actually, classes 3 and 11 have no sub-groups at all.

Table 8: SAMPLES OF SIX NEUTER NOUN DECLENSION PATTERNS

Number	Number On Chart	Number of Declension Pattern	Sample Word	Animate or Inanimate
1	8	3	mesto (no sub-group)	i
2	9	4a	pole	i
3	10	4b	dejstvie	i
4	11	4c	lozhe	i
5	12	4d	plat6e	i
6	30	11	ukho (2 stems, no sub-group)	i

---

\*There are some twenty neuter nouns which are so irregular that they will have to be declined manually; e.g., vremia, vymia, ditia, imia, plamia, plemia, semia, stremia, temia, znamia.

N-3 Class 3 is one of two (class 11) which has no sub-groups.

Table 9: CLASS 3

C	S	P
	3(1)	3(1)
1	o	a
2	a	ø
3	u	am
4	o	a
5	om	ami
6	e	akh

Samples: mesto, slovo, delo, ozero, selo

N-4 Class 4 includes inanimate neuter nouns with four slightly different declension patterns.

Table 10: CLASS 4

C	S				P			
	4a(i)	4b(i)	4c(i)	4d(i)	4a(i)	4b(i)	4c(i)	4d(i)
1	e	e	e	e	ia	ia	a	ia
2	ia	ia	a	ia	ej	j	ø	ev
3	iu	iu	u	iu	iam	iam	am	iam
4	e	e	e	e	ia	ia	a	ia
5	em	em	em	em	iami	iami	ami	iami
6	e	i	e	e	iakh	iakh	akh	iakh

Samples: 4a: pole, more  
 4b: dejstvie  
 4c: lozhe, zhilisce, uchilisce, serdtse  
 4d: plat6e

N-11 Class 11 is one of two (class 3) which has no sub-groups.

Table 11: CLASS 11

C	S	P
	11(i)	11(i)
1	o	i
2	a	ej
3	u	am
4	o	i
5	om	ami
6	e	akh



E. Declension of Feminine Nouns

There are three main classes or seven groups of feminine declension patterns for nouns. Like the neuter nouns, the feminine nouns are much more regular than those of the masculine gender.\* The masculine sub-groups 5c and 6d have a feminine declension type, but they have been discussed at the end of the section on masculine declensions.

Table 12: SAMPLES OF SEVEN FEMININE NOUN DECLENSION PATTERNS

Number	Number On Chart	Number of Declension Pattern	Sample Word	Animate or Inanimate
1	13	5a	bukva	b
2	14	5b	kniga	i
3	16	6a	dolia	b
4	17	6b	funktsiia	i
5	18	6c	niania	a
6	20	7a	chast6	i
7	21	7b	loshad6	a

\* A few feminine nouns are irregular, but because of their small number it is not worthwhile to establish a special sub-group for them. They will have to be declined manually; e.g., zmeia, gost6ia, and other animate nouns which somewhat differ from type 6b; or belil6nia, baryshnia, kolokol6nia, bojnia, derevnia, etc., which in the genitive plural end in -en and kukhnia which ends in -on.

F-5 The patterns of classes 5a and 6a apply to both animate and inanimate nouns. Class 5a includes both animate and inanimate nouns; 5b applies to inanimate nouns only, while 5c is a masculine sub-group of animate nouns with a feminine declension pattern (cf. 6d: both are described among the masculine nouns).

Table 13: CLASS 5

C	S			P		
	fem. 5a(b)	fem. 5b(i)	masc. 5c(a)	fem. 5a(b)	fem. 5b(i)	masc. 5c(a)
1	a	a	a	y	i	i
2	y	i	i	ø	ø	ej
3	e	e	e	am	am	am
4	u	u	iu	y	i	ej
5	oj	oj/oiu	ej	ami	ami	ami
6	e	e	e	akh	akh	akh

Samples, 5a: bukva, zhena, trava

5b: kniga, tysiacha, devushka, ruchka (2 stems),  
palka (2 stems)

5c: muzhchina, dedushka, mal6chishka

F-6 Class 6a includes both animate and inanimate nouns; 6b contains only inanimate and 6c animate nouns, while 6d is a masculine sub-group with a feminine declension pattern (cf. 5c; both are described among the masculine nouns).

Table 14: CLASS 6

C	S				P			
	fem. 6a(b)	fem. 6b(i)	fem. 6c(a)	masc. 6d(a)	fem. 6a(b)	fem. 6b(i)	fem. 6c(a)	masc. 6d(a)
1	ia	ia	ia	ia	i	i	i	i
2	i	i	i	i	ej	j	ej	ej
3	e	i	e	e	iam	iam	iam	iam
4	iu	iu	iu	iu	i	i	ej	ej
5	ej	ei	ej	ej	iami	iami	iami	iami
6	e	i	e	e	iakh	iakh	iakh	iakh

Samples, 6a: barynia, svinia, stat6ia (2 stems)

6b: ideia, istoria, funktsia, armia

6c: niania (2 stems), pulia (2 stems)

6d: sud6ia, diadia

F-7 This class contains a group of inanimate and another for animate nouns.

Table 15: CLASS 7

C	S		P	
	7a(i)	7b(a)	7a(i)	7b(a)
1	6	6	i	i
2	i	i	ej	ej
3	i	i	iam	iam
4	6	6	i	ej
5	6iu	6iu	iam1	iam1
6	i	i	iakh	iakh

Samples, 7a: chast6, gavan6, step6, artel6

7b: loshad6

F. Russian Suffixes in English Transliteration

The morphology of Russian nouns is limited to 26 suffixes. These are arranged according to the number of letters in their English transliteration and alphabetically within those groups. They are preceded by the symbols Ø and 6.

Table 16: NUMBER OF NOUN SUFFIXES

Group Number	Number of English Letters	Number of Suffixes
1	1	9
2	2	9
3	3	6
4	4	2
Total		26

The suffixes have been given numeric octal notations in our working chart ranging from 001 to 202 (table 17, col. 1 and table 18, col. 2). The noun stems have been divided into 30 classes ranging from 100 to 130\*. In accordance with these code pattern numbers, the computer will select the appropriate suffixes for every stem entered and compile a complete set of paradigmatic forms for every word desired.

---

\* Code pattern no. 100 pertains to undeclinable nouns. The code pattern numbers 100 to 130 for the noun stems will have to be inserted manually.

Table 17: NOUN SUFFIXES

Declension Class	Suffix	Noun Classes
001	ø (0, zero)	100, 101, 102, 106, 108, 111, 113, 114, 122, 123, 126, 127, 128, 129
002	6	104, 105, 118, 120, 121, 124, 125
003	a	101, 102, 106, 108, 111, 113, 114, 115, 122, 123, 126, 127, 128, 129, 130
004	e	101, 102, 103, 104, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 118, 119, 122, 123, 124, 125, 126, 127, 128, 129, 130
005	i	102, 103, 104, 105, 106, 107, 110, 114, 115, 116, 117, 118, 119, 120, 121, 122, 124, 127, 128, 130
006	j	103, 107, 110, 117
011	o	108, 130
012	u	101, 102, 106, 108, 111, 113, 114, 115, 122, 123, 126, 127, 128, 129, 130
014	y	101, 113, 123
022	am	101, 102, 106, 108, 111, 113, 114, 115, 122, 123, 126, 127, 128, 129, 130
027	ej	104, 105, 106, 109, 115, 116, 118, 119, 120, 121, 124, 125, 127, 128, 130
030	em	103, 104, 105, 106, 107, 109, 110, 111, 112, 124, 125, 128
033	ev	103, 107, 112
034	ia	103, 104, 107, 109, 110, 112, 116, 117, 118, 119, 124, 125
043	iu	103, 104, 107, 109, 110, 112, 116, 117, 118, 119, 124, 125
054	oj	113
055	om	101, 102, 108, 122, 123, 126, 127, 129, 130
056	ov	101, 102, 122, 123, 126, 129

Approved For Release 2004/01/15 : CIA-RDP64-00046R000200030003-3

Declension Class	Suffix	Noun Classes
101	6iu	120, 121
103	akh	101, 102, 106, 108, 111, 113, 114, 115, 122, 123, 126, 127, 128, 129, 130
104	ami	101, 102, 106, 108, 111, 113, 114, 115, 122, 123, 126, 127, 128, 129, 130
112	eiū	117
123	iam	103, 104, 105, 107, 109, 110, 112, 116, 117, 118, 119, 120, 121, 124, 125
140	oiu	114
201	iakh	103, 104, 105, 107, 109, 110, 112, 116, 117, 118, 119, 120, 121, 124, 125
202	iami	103, 104, 105, 107, 109, 110, 112, 116, 117, 118, 119, 120, 121, 124, 125

Approved For Release 2004/01/15 : CIA-RDP64-00046R000200030003-3

Table 18: SUFFIX CODE NUMBERS FOR AUTOMATIC DECLENSION OF NOUNS

Number of Noun Stems	Suffix Code Numbers for Noun Stems
100	001
101	001, 003, 004, 012, 014, 022, 055, 056, 103, 104
102	001, 003, 004, 005, 012, 022, 055, 056, 103, 104
103	004, 005, 006, 030, 033, 034, 043, 123, 201, 202
104	002, 004, 005, 027, 030, 043, 123, 201, 202
105	002, 005, 027, 030, 123, 201, 202
106	001, 003, 004, 005, 012, 022, 027, 030, 103, 104
107	004, 005, 006, 030, 033, 034, 043, 123, 201, 202
108	001, 003, 004, 011, 012, 022, 055, 103, 104
109	004, 027, 030, 034, 043, 123, 201, 202
110	004, 005, 006, 030, 034, 043, 123, 201, 202
111	001, 003, 004, 012, 022, 030, 103, 104
112	004, 030, 033, 034, 043, 123, 201, 202
113	001, 003, 004, 012, 014, 022, 054, 103, 104
114	001, 003, 004, 005, 012, 014, 022, 103, 104, 140
115	003, 004, 005, 012, 022, 027, 103, 104
116	005, 006, 034, 043, 112, 123, 201, 202
117	005, 006, 034, 043, 112, 123, 201, 202
118	002, 004, 005, 027, 034, 043, 123, 201, 202
119	004, 005, 027, 034, 043, 123, 201, 202
120	002, 005, 027, 101, 123, 201, 202
121	002, 005, 027, 101, 123, 201, 202
122	001, 003, 005, 012, 022, 055, 056, 103, 104
123	001, 003, 004, 012, 014, 022, 055, 056, 103, 104



Number of Noun Stems	Suffix Code Numbers for Noun Stems
124	002, 004, 005, 027, 030, 034, 043, 123, 201, 202
125	002, 004, 027, 030, 034, 043, 123, 201, 202
126	001, 003, 004, 012, 022, 055, 056, 103, 104
127	001, 003, 004, 005, 012, 022, 027, 055, 103, 104
128	001, 003, 004, 005, 012, 022, 027, 030, 103, 104
129	001, 003, 004, 012, 022, 055, 056, 103, 104
130	003, 004, 005, 011, 012, 022, 027, 055, 103, 104

## III. SAMPLE DECLENSION

Based on the declension pattern which will be inserted manually, the computer will select the appropriate suffix codes and compile automatically a complete set of paradigmatic forms for every word desired. For instance, the stems of the two fem. nouns:

knig- and tysiach-

Table 19. Declension of fem. Nouns of Class 5b

Case	Suffix Code	Suffix
P-2 and stem	001	Ø
S-1	003	a
S-3, 6	004	e
S-2, P-1, 4	012	i
S-4	014	u
P-3	022	am
S-5*	054	oj
P-6	103	akh
P-5	104	ami
S-5*	112	oiu

\* The machine will identify both forms of the inst. sing., the modern form knig-oj and the obsolescent form knig-oiu.

Table 20: SAMPLE DECLENSION OF A TWO-STEM NOUN

Sing. Case	Sing. Stem	Suffix Code Number	Suffix	Plur. Case	Plur. Stem	Suffix Code Number	Suffix
2	ukh-	003	a	1, 4	ush-	005	i
6	ukh-	004	e	3	ush-	022	am
1, 4	ukh-	011	o	2	ush-	027	ej
3	ukh-	012	i	6	ush-	103	akh
5	ukh-	055	om	5	ush-	104	ami

In this particular case, the stems are equally divided between the singular and the plural. It is, however, quite irrelevant for the system if the two stems are used for singular and/or plural.

#### IV. TWO-STEM NOUNS

We encounter two-stem nouns only in combination with the following five suffixes:

Table 21: SUFFIXES OF TWO-STEM NOUNS

Suffix	Suffix Code
Ø	001
6	002
a	003
o	011
ia	034

On the basis of the two stems the machine produces an unambiguous and correct declension.

Three-stem nouns belong to the 5% irregular nouns which will be declined manually. We are only interested in providing for the declension of the nouns

## V. CONCLUDING REMARK

This system of dictionary preparation was developed especially for the Unified Transfer System. It applies, however, equally to the compilation of any other dictionary. The present report is limited to Russian nouns, but a corresponding analysis has been drafted for the declension of remaining nominals (numerals, etc.), as well as all modifiers (adjectives, numerals, long forms of participles, etc.), and the conjugation of verbs. The three reports will eventually be incorporated in a working manual.

CHART 1: DECLENSION OF RUSSIAN NOUNS

Gender	Declen- sion Pattern Number	Declen- sion Type	Animate or In- animate	Singular						Plural						Samples	Remarks
				1	2	3	4	5	6	1	2	3	4	5	6		
various	100	none	both														
masc.	101	1a	inan.	g	a	u	g	om	e	y	ov	am	y	ani	akh	doklad	some foreign loan words and certain foreign names
masc.	102	1b	inan.	g	a	u	g	om	e	i	ov	am	i	ani	akh	ocherk	very large group; for animate nouns, see 8b and 10a-b
masc.	103	2a	inan.	j	ia	iu	j	em	e	i	ev	iam	i	iam	iakh	saraj	close to types 8a, 10a-b
masc.	104	2b	inan.	6	ia	iu	6	em	e	i	ej	iam	i	iam	iakh	kamen6	corresponds to 2e (anim.)
masc.	105	2c	inan.	6	i	i	6	em	i	i	ej	iam	i	iam	iakh	put6	very close to 10b (inan.)
masc.	106	2d	inan.	g	a	u	g	em	e	i	ej	am	i	ani	akh	ehkipazh	corresponds to 9a (anim.)
masc.	107	2e	anim.	j	ia	iu	ia	em	e	i	ev	iam	ev	iam	iakh	geroj	corresponds to 2a (anim.)
neut.	108	3	inan.	o	a	u	o	om	e	a	g	am	a	ani	akh	mesto	no sub-group (see also 11)
neut.	109	4a	inan.	e	ia	iu	e	em	e	ia	ej	iam	ia	iam	iakh	pole	
neut.	110	4b	inan.	e	ia	iu	e	em	i	ia	j	iam	ia	iam	iakh	dejstvie	
neut.	111	4c	inan.	e	a	u	e	em	e	a	g	am	a	ani	akh	lozhe	
neut.	112	4d	inan.	e	ia	iu	e	em	e	ia	ov	iam	ia	iam	iakh	plat6e	
fem.	113	5a	both	a	y	e	u	oj	e	y	g	am	y	ani	akh	bukva	
fem.	114	5b	inan.	a	i	e	u	oj (olu)	e	i	g	am	i	ani	akh	kniga	
masc.	115	5c	anim.	a	i	e	u	ej	e	i	ej	am	ej	ani	akh	muzhchina	small masc. sub-group with fem. declension type (see also 6d)
fem.	116	6a	both	ia	i	e	iu	ej	e	i	ej	iam	i	iam	iakh	svin6ia	
fem.	117	6b	inan.	ia	i	i	iu	ei	i	i	j	iam	i	iam	iakh	funktsia	irreg. fem. nouns like zmeia and gost6ia to be declined manually
fem.	118	6c	anim.	ia	i	e	iu	ej	e	i	6	iam	6	iam	iakh	niania	
masc.	119	6d	anim.	ia	i	e	iu	ej	e	i	ej	iam	ej	iam	iakh	sud6ia	small masc. sub-group with fem. declension type (see also 5c)
fem.	120	7a	inan.	6	i	i	6	6iu	i	i	ej	iam	i	iam	iakh	chast6	
fem.	121	7b	anim.	6	i	i	6	6iu	i	i	ej	iam	ej	iam	iakh	loshad6	gen. and acc. plur. are identical
masc.	122	8a	anim.	g	a	u	a	om	e	i	ov	am	ov	ani	akh	letchik	close to type 1b
masc.	123	8b	anim.	g	a	u	a	om	e	y	ov	am	ov	ani	akh	lev (2 stems)	corresponds to 1a (inan.)
masc.	124	9a	anim.	6	ia	iu	ia	em	e	i	ej	iam	ej	iam	iakh	kon6	corresponds to 2b (inan.)
masc.	125	9b	anim.	6	ia	iu	ia	em	e	ia	ej	iam	ej	iam	iakh	uchitel6	see also 10d
masc.	126	10a	inan.	g	a	u	g	om	e	a	ov	am	a	ani	akh	gorod	close to la-b; corresponds to 10d (anim.)
masc.	127	10b	inan.	g	a	u	g	om	e	i	ej	am	i	ani	akh	nozh	close to la-b; see also 8a-b (anim.)
masc.	128	10c	anim.	g	a	u	a	em	e	i	ej	am	ej	ani	akh	tovarisc	
masc.	129	10d	anim.	g	a	u	a	om	e	a	ov	am	ov	ani	akh	doktor	see 9b (inan.)
neut.	130	11	inan.	o	a	u	o	om	e	i	ej	am	i	ani	akh	ukho (2 stems)	no sub-group (see also type 3)

SECTION THREE

REPORT ON COMPUTER IMPLEMENTATION  
OF THE UNIFIED TRANSFER SYSTEM

By

B. D. Blickstein

The basic flow diagram, Figure 1 on the next page, traces the basic functions which the computer must follow, and shows the necessary magnetic tape configuration. Also shown on this chart is an index to the tapes, showing the processes in which each tape is involved.

The flow chart is divided into the following computer program steps:

1. Text Preparation

The entry to this box is the raw text, prepared by either key-punching from the Russian or by a character-scanning device. The function of this program is to convert the text to a form which the machine may more easily accept. At the same time, Romanized expressions will be extracted and saved for later re-entry into the system. At this point, a transliteration of the text can be produced.

2. Alpha Sort

The sequenced and prepared text is now sorted into dictionary order. The original text sequence numbers are retained.

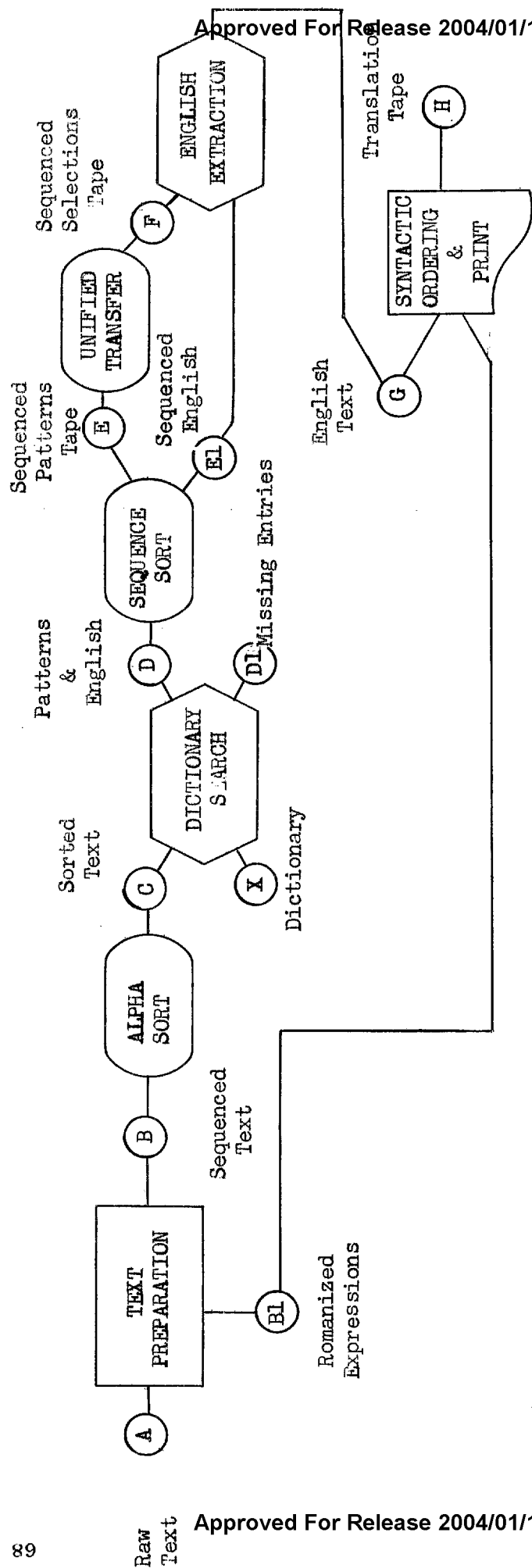
3. Dictionary Search

The sorted text is matched against the dictionary tape. For each text entry for which a dictionary match exists, a record will be written on tape D, consisting of the appropriate pattern number and the set of English meanings, still retaining the text sequence number. For each text entry which has no match, a dummy "word missing" record will be written, and the Russian word written on the "missing entries" tape D1 for subsequent printing.

4. Sequence Sort

Tape D is now sorted back into text sequence. At the end of the sort, a split of the tape D record will occur, creating two tapes, E and E1; Tape E contains only pattern numbers, and tape E1 the corresponding sets of English meanings.

# THE UNIFIED TRANSFER SYSTEM BASIC FLOW DIAGRAM



TAPE	NAME	OUTPUT FROM:	INPUT TO:	FORMAT
A	Raw Text Input	Text Preparation	Text Preparation	BCD
B	Sequenced Text	Text Preparation	Alpha Sort	Binary
BL	Romanized Expressions	Text Preparation	Syntactic Ordering	Binary
C	Sorted Text	Alpha Sort	Dictionary Search	Binary
D	Patterns & English	Dictionary Search	Sequence Sort	Binary
DL	Missing Entries	Dictionary Search	Sequence Sort	BCD
E	Sequenced Patterns	Sequence Sort	Unified Transfer	Binary
EL	Sequenced English	Sequence Sort	English Extraction	Binary
F	Sequenced Selections	Unified Transfer	English Extraction	Binary
G	English Text	English Extraction	Syntactic Ordering	Binary
H	Translation Output	Syntactic Ordering	Dictionary Search	BCD
X	Dictionary	Syntactic Ordering	Dictionary Search	Binary



#### 5. Unified Transfer

The basic code matching algorithms are performed here. Blocks are recognized, and the proper meaning selections are made, the output is the sequenced selections tape. The computer considerations of this section will be treated later at some length.

#### 6. English Extraction

The selections tape is used to select the proper English meaning from the English tape at this point. The output is an English text with certain block marks present.

#### 7. Syntactic Ordering

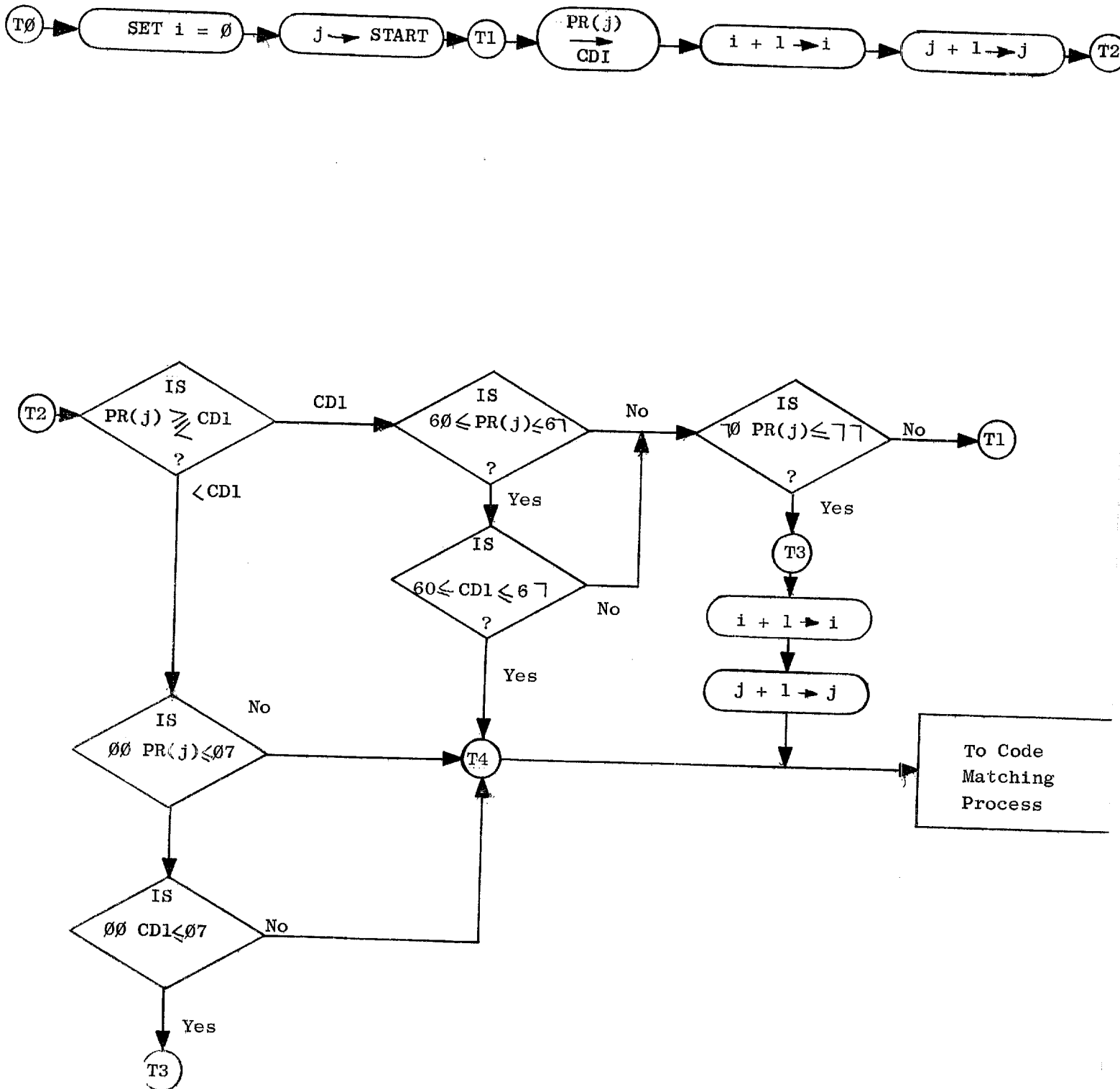
Re-arrangement of the syntactic blocks is performed here; at the same time, the Romanized expressions are merged back into the text, and a final translation tape, suitable for printing, is produced.

Some discussion of the matching algorithms is appropriate here; the first part of the process is shown in Figure 2 on the next page. This involves the identification of phrases by means of the parts-of-speech code numbers, which we shall refer to as progression numbers. Let  $PR(j)$  be the progression number associated with the  $j^{\text{th}}$  text sequence. As the translation progresses, suppose all phrases through the  $(j-1)^{\text{th}}$  are strung, and we thus wish to find the boundaries of the phrase beginning with this  $j^{\text{th}}$  word. The flow chart (beginning at step  $t_0$ ) traces the entire technique for identifying the phrase. At the conclusion of this process, the phrase is bounded, and the code matching on the actual dictionary patterns may commence. It can be seen that this algorithm involves little else than a few arithmetic counts and comparisons, and certainly no analysis of the source language is performed. This example serves well to point up the essential philosophy of the Unified Transfer technique; the computer is used for the things it does best, namely arithmetic and logic, while the analysis is done in advance by means of the dictionary. We do not ask the computer

Fig. 2

## Unified Transfer

## Phrase Identification



to come to conclusions about form; we merely ask it to choose between various possible forms on a basis of simple logical rules. In this way, the full power of the machine is used in the most efficient manner.

The subsequent code matching process is also designed with this same philosophy. The only question asked is basically an "equal-or-unequal" choice; blocking for syntactic re-arrangement is similarly well suited to this type of treatment. In no case does the machine ever "know" about syntax or meaning; it only follows completely abstract rules for operating on certain numerical sequences.